

Implementasi Algoritma C4.5 untuk Analisa Performa Pelayanan Bank Terhadap Nasabah

Implementation of C4.5 Algorithm for Bank Service Performance Analysis To Customer

Fandi Ramadhan Wiratama¹, Setia Astuti²

^{1,2}Jurusan Teknik Informatika, Universitas Dian Nuswantoro Semarang

Jl. Imam Bonjol 205-207 Semarang 50131

E-mail : fandiramadhanwiratama@gmail.com¹, setia.astuti@dsn.dinus.ac.id²

Abstrak

Perbankan adalah segala sesuatu yang berkaitan dengan bank, kegiatan usaha, serta cara dan proses dalam melaksanakan kegiatan usahanya. Tujuan dari penelitian ini adalah untuk mengetahui tingkat kepuasan nasabah terhadap pelayanan yang diberikan oleh Bank Danamon Kantor Cabang Utama Semarang dan mengetahui tingkat akurasi untuk prediksi kepuasan pelanggan. Kualitas layanan yang ingin dipenuhi harus dilihat dari sudut pandang nasabah, karena nasabah bank merupakan penilai utama dari kualitas pelayanan sehingga prioritas utama dalam jaminan kualitas adalah penilaian nasabah terhadap kualitas pelayanan jasa perbankan. Metode yang digunakan dalam penelitian ini menggunakan metode algoritma C4.5. Algoritma C4.5 adalah salah satu metode algoritma yang digunakan oleh beberapa penelitian guna membuat suatu pohon keputusan. Pohon keputusan merupakan metode klasifikasi serta prediksi yang kuat dan terkenal. Metode menggunakan pohon keputusan merupakan metode yang dapat mengubah suatu fakta yang besar menjadi suatu pohon keputusan yang merepresentasikan aturan agar mudah dipahami dengan bahasa alami. Atribut yang digunakan dalam metode ini yaitu pelayanan, fasilitas, dan mobile banking. Kesimpulan yang didapatkan dalam penelitian ini menunjukkan bahwa nilai akurasi pada pengujian ke-2 dapat dikatakan baik karena memiliki nilai akurasi lebih tinggi di bandingkan pengujian 1.

Kata kunci : kepuasan pelanggan, C4.5, data mining, pohon keputusan

Abstract

Banking is everything related to bank and business activities, as well as the manner and process of carrying out its business activities. Quality of service is to be met to be seen from the perspective of the customer, because the customer bank is the main evaluator of quality of service so that the top priority in quality assurance is the assessment of our customers for quality banking services. The method used in this study using C4.5 algorithm. C4.5 algorithm is one method algorithm used by some research to make a decision tree. The decision tree is a method of classification and prediction of strong and famous. The method of using a decision tree is a method that can change the fact that big into a decision tree that represents the rule to be easily understood by the natural language. Attribute used in this method are services, facilities, and mobile banking. Those results indicate that the value of accuracy in all the second test can be said as good because it has a higher accuracy in comparison with test 1.

Keywords : customer's satisfaction, C4.5, data mining, desicion tree.

1. Pendahuluan

Perbankan adalah segala sesuatu yang berkaitan dengan bank, kegiatan usaha, serta cara dan proses dalam melaksanakan kegiatan usahanya. Bank adalah lembaga keuangan yang kegiatannya menghimpun dan menyalurkan dana dari dan kepada masyarakat yang memiliki fungsi memperlancar lalu lintas pembayaran, dengan kata lain bank adalah salah satu lembaga keuangan yang usaha pokoknya memberikan kredit serta jasa dalam lalu lintas pembayaran dan peredaran uang [1]. Lembaga keuangan perbankan memiliki peranan penting dalam perekonomian suatu Negara. Bisnis perbankan merupakan bisnis kepercayaan, dimana bank harus mampu memberikan rasa aman pada nasabahnya menyangkut dana yang disimpan atau yang telah dikelola oleh bank tersebut [2].

Kualitas layanan yang ingin dipenuhi harus dilihat dari sudut pandang nasabah, karena nasabah bank merupakan penilai utama dari kualitas pelayanan sehingga prioritas utama dalam jaminan kualitas adalah penilaian nasabah terhadap kualitas pelayanan jasa perbankan.

Dalam penelitian ini, peneliti mengambil metode Algoritma C4.5. Algoritma C4.5 adalah salah satu metode algoritma yang digunakan oleh beberapa penelitian guna membuat suatu pohon keputusan [3]. Pohon keputusan merupakan metode klasifikasi serta prediksi yang kuat dan terkenal [4]. Metode menggunakan pohon keputusan merupakan metode yang dapat mengubah suatu fakta yang besar menjadi suatu pohon keputusan yang merepresentasikan aturan agar mudah dipahami dengan bahasa alami [5, 6, 7].

2. Metode Penelitian

2.1. Metode Pengumpulan Data

Metode yang digunakan dalam penelitian ini adalah menyebar kuesioner. Kuesioner merupakan teknik pengumpulan data yang dilakukan dengan cara memberikan seperangkat pertanyaan atau pernyataan kepada orang lain yang dijadikan respon dan untuk dijawabnya.

2.2. Evaluasi dan Validasi

Setelah data diuji menggunakan Matlab, maka akan terbentuk pohon keputusan dari data tersebut. Kemudian akan dilakukan penghitungan akurasi, *precision*, *recall* dan *f-measure*, selanjutnya dilakukan evaluasi dan validasi hasil dihitung menggunakan rumus akurasi, *precision recall* dan *f-measure* berikut ini [8]:

- a. Menguji metode/ model menggunakan *data testing*.

Rumus untuk menghitung akurasi adalah sebagai berikut:

$$\text{Akurasi} = \frac{\text{Jumlah yang diklasifikasi benar}}{\text{Total sample testing yang diuji}} \quad (1)$$

- b. Menghitung *precision* dan *recall* untuk mengetahui akurasinya.

Rumus *Precision* dan *Recall* adalah sebagai berikut:

$$p = \frac{TP}{TP + FP} \quad r = \frac{TP}{TP + FN} \quad (2)$$

Keterangan:

p = *Precision*

r = *Recall*

TP = *True Positive*

FP = *False Positive*

FN = *False Negative*

- c. Menghitung *F-Measure*.

F-Measure merupakan *harmonic mean* dari *precision* dan *recall*. Berikut rumus untuk menghitung *F-Measure* :

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

Keterangan :

F1 = *F-Measure*

2.3. Teknik Analisis Data

Data yang digunakan dalam penelitian ini adalah data berdasarkan kuisisioner yang diberikan secara langsung pada Nasabah Bank Danamon Kantor Cabang Utama. Metode yang disulkan untuk proses seperti yang telah dijelaskan di atas yaitu dengan menggunakan metode Algoritma C4.5.

3. Hasil dan Analisis

3.1. Pengujian Rules Terhadap Data Pelanggan

Pengujian terhadap validitas sistem bertujuan untuk mengetahui apakah solusi yang dihasilkan oleh pohon keputusan tersebut *valid* atau tidak. *Rules* tersebut dikatakan *valid* jika jumlah dan nama pelanggan yang puas sesuai dengan *data set*.

Data set dibagi menjadi dua bagian, yaitu *data training* dan *data testing*, yaitu:

- Data *training* 60% dan data *testing* 40%
- Data *training* 80% dan data *testing* 20%

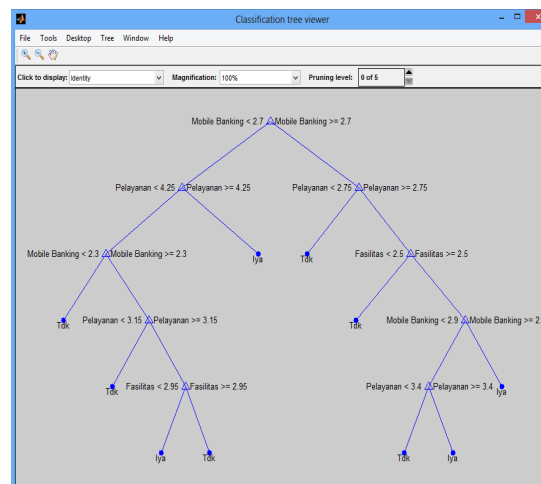
1. Pengujian Pertama

Data *set* dibagi menjadi dua 60% data *training* dan 40% data *testing*. Keseluruhan data *set* berjumlah 150, data *training* sebanyak 90 data dan data *testing* sebanyak 60 data. Seperti pada Tabel 1:

Tabel 1. Data *training* 60%

Nama Pelanggan	Pelayanan	Fasilitas	Mobile Banking	Hasil
1	2.7	2.9	2.8	Tidak
2	3.7	3.2	2.8	Ya
...
89	3.7	3.3	3	Ya
90	3.3	3.3	3	Ya

Tabel 1 merupakan tabel data pelanggan yang akan di *training* menggunakan *Software* Matlab. Data *training* yang berjumlah 90 data pelanggan. Kemudian data tersebut dicopy di notepad dan kemudian dibuat pohon keputusan. Seperti pada Gambar 3.



Gambar 3. Pohon Keputusan Data *Training* 60%

Pohon keputusan diatas menghasilkan *rules* sebagai berikut:

Rules untuk Hasil = Tdk	
1.	IF Mobile Banking<2.7 and Pelayanan<4.25 and Mobile Banking<2.3 THEN Hasil = Tdk
2.	IF Mobile Banking<2.7 and Pelayanan<4.25 and Mobile Banking>=2.3 and Pelayanan <3.15 THEN Hasil = Tdk
3.	IF Mobile Banking<2.7 and Pelayanan<4.25 and Mobile Banking >=2.3 and Pelayanan >=3.15 and Fasilitas >=2.95 THEN Hasil = Tdk
4.	IF Mobile Banking>=2.7 and Pelayanan <2.75 THEN Hasil = Tdk
5.	IF Mobile Banking>=2.75 and Pelayanan >=2.75 and Fasilitas<2.5 THEN Hasil = Tdk
6.	Mobile Banking>=2.7 and Pelayanan>=2.25 and Fasilitas >=2.5 and Mobile Banking <2.9 and Pelayanan <3.4 THEN Hasil = Tdk

Rules untuk Hasil = Iya	
1.	IF Mobile Banking<2.7 and Pelayanan<4.25 and Mobile Banking>=2.3 and Pelayanan>=3.15 and Fasilitas <2.95 THEN Hasil = Iya
2.	IF Mobile Banking<2.7 and Pelayanan >=4.25 THEN Hasil = Iya
3.	IF Mobile Banking>=2.7 and Pelayanan >=2.75 and Fasilitas>=2.5 and Mobile Banking<2.9 and Pelayanan >=3.4 THEN Hasil = Iya
4.	IF Mobile Banking >=2.7 and Pelayanan >=2.75 and Fasilitas 2.5 and Mobile Banking >=2.9 THEN Hasil = Iya

Setelah dilakukan training terhadap 90 data diatas maka selanjutnya dilakukan an data *testing* terhadap data *set* yang tersisa yaitu 60 data pelanggan. Seperti pada Tabel 2:

Tabel 2. Data *Testing* 40%

Nama Pelanggan	Pelayanan	Fasilitas	Mobile Banking	Hasil	Prediksi
91	3.5	2.7	3	Ya	Ya
92	3.5	3.3	3.4	Ya	Ya
...
149	3	2.9	2.8	Ya	Tidak
150	3	2.7	2.8	Ya	Tidak

Pada Tabel 2 menunjukkan bahwa ada 13 data yang hasil prediksinya tidak tepat, ditunjukan dengan kolom yang berwarna biru. Dapat dilihat dari Tabel 3.

Tabel 3.Hasil *Testing* Pertama

Jumlah Data <i>testing</i>	Erro r	Positif Benar (TP)	Positif Salah (FP)	Negatif Salah (FN)
60	13	42	9	5

Pada tabel di atas data dengan perbandingan 60% : 40% terdapat *error* sebanyak 13 data, positif benar 42, positif salah 9 dan negatif salah 5.

Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasi secara benar dengan total sample *data testing* yang diuji.

$$\begin{aligned} \text{Akurasi} &= \frac{\text{Data Testing} - \text{Error}}{\text{Jumlah Data testing}} \\ &= \frac{60 - 13}{60} \\ &= 0.78333333 \\ &= 78\% \end{aligned}$$

Precision merupakan jumlah sample berkategori positif diklasifikasi benar dibagi dengan total sampel yang diklasifikasi sebagai sample positif. Menghitung nilai *precision* dihitung dengan cara membagi jumlah data benar yang bernilai positif (*True Positive*) dibagi dengan jumlah data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai positif (*False Negative*).

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \\ &= \frac{42}{42 + 9} \\ &= 0.82352941 \\ &= 82\% \end{aligned}$$

Recall merupakan jumlah sampel diklasifikasi positif dibagi total sampel dalam testing set berkategori positif. *Recall* dihitung dengan cara membagi data benar yang bernilai positive (*True Positive*) dengan hasil penjumlahan dari data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai negatif (*False Negative*).

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN} \\ &= \frac{42}{42 + 4} \\ &= 0.91304348 \\ &= 91\% \end{aligned}$$

F-Measure merupakan *harmonic mean* dari *precision* dan *recall*. Nilai *F-Measure* didapat dari perhitungan pembagian hasil dari perkalian *precision* dan *recall* dengan hasil penjumlahan *precision* dan *recall*, kemudian dikalikan dua.

$$\begin{aligned} \text{F-Measure} &= \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \\ &= 0.86597938 \\ &= 87\% \end{aligned}$$

Tabel 4.Evaluasi dan Validasi Pertama

Presentasi Data	Data Training	Data Testing	Akurasi	Precision	Recall	F-Measure
60% : 40%	90	60	78%	82%	91%	87%

Pada tabel 4 data dengan perbandingan 60% : 40% memiliki nilai akurasi sebesar 78%, *recall* sebesar 91%, *precision* 82% dan *f-measure* 87%.

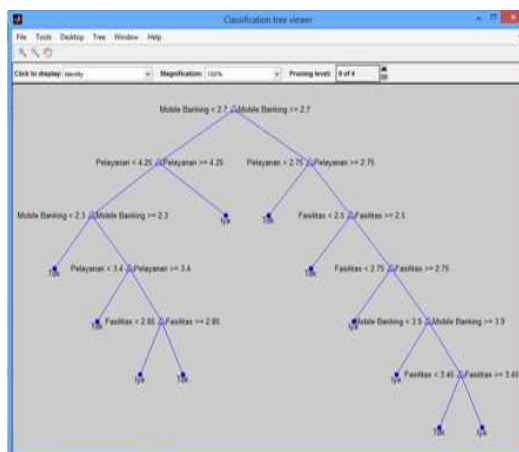
2. Pengujian Kedua

Data set dibagi menjadi dua 80% data *training* dan 20% data *testing*. Keseluruhan data set berjumlah 150, data *training* sebanyak 120 data dan data *testing* sebanyak 30 data.

Tabel 5. DataTraining 80%

Nama Pelanggan	Pelayanan	Fasilitas	Mobile Banking	Hasil
1	2.7	2.9	2.8	Tidak
2	3.7	3.2	2.8	Ya
...
119	3.7	3.1	3.2	Ya
120	2.7	2.9	2.6	Tidak

Tabel 5 merupakan tabel data pelanggan yang akan di *training* menggunakan *Software* Matlab. Data training yang berjumlah 120 data pelanggan. Kemudian data tersebut dicopy di notepad dan kemudian di buat pohon keputusan. Seperti pada Gambar 4.

**Gambar 4.** Pohon Keputusan Data Training 80%

Pohon keputusan diatas menghasilkan *rules* sebagai berikut:

Rules untuk Hasil = Tdk	
1.	IF Mobile Banking <2.7 and Pelayanan <4.25 and and Mobile Banking <2.3 THEN Hasil = Tdk
2.	IF Mobile Banking <2.7 and Pelayanan <4.25 and Mobile Banking >=2.3 and Pelayanan <3.4 THEN Hasil = Tdk
3.	IF Mobile Banking <2.7 and Pelayanan <4.25 and Mobile Banking >=2.3 and Pelayanan >=3.4 and Fasilitas >=2.85 THEN Hasil = Tdk
4.	IF Mobile Banking >=2.7 and Pelayanan <2.75 THEN Hasil = Tdk
5.	IF Mobile Banking >=2.7 and Pelayanan >=2.75 and Fasilitas <2.5 THEN Hasil = Tdk
6.	IF Mobile Banking >=2.7 and Pelayanan >=2.75 and Fasilitas >=2.5 and Fasilitas >=2.75 and Mobile Banking >=3.9 and Fasilitas <3.45 THEN Hasil = Tdk

Rules untuk Hasil = Iya	
1. IF Mobile Banking <2.7 and Pelayanan <4.25 and Mobile Banking >=2.3 and Pelayanan >=3.4 and Fasilitas <2.85 THEN Hasil = Iya	
2. IF Mobile Banking <2.7 and Pelayanan >=4.25 THEN Hasil = Iya	
3. IF Mobile Banking >=2.7 and Pelayanan >=2.75 and Fasilitas >=2.5 and Fasilitas <2.75 THEN Hasil = Iya	
4. IF Mobile Banking >=2.7 and Pelayanan >=2.75 and Fasilitas >=2.5 and Fasilitas >=2.75 and Mobile Banking <3.9 THEN Hasil = Iya	
5. IF Mobile Banking >=2.7 and Pelayanan >=2.75 and Fasilitas >=2.5 >=2.75 and Mobile Banking >=3.9 and Fasilitas >=3.45 THEN Hasil = Iya	

Setelah dilakukan *training* terhadap 120 data diatas maka selanjutnya di lakukan data *testing* terhadap data *set* yang tersisa yaitu 30 data pelanggan seperti pada Tabel 5:

Tabel 6. Data *Testing* 20%

Nama Pelanggan	Pelayanan	Fasilitas	Mobile Banking	Hasil	Prediksi
121	3	3.6	3.4	Ya	Ya
122	3	2.8	2.4	Tidak	Tidak
...
149	3	2.9	2.8	Ya	Ya
150	3	2.7	2.8	Ya	Ya

Pada Tabel 5 menunjukkan bahwa ada 5 data yang hasil prediksinya tidak tepat, ditunjukan dengan kolom yang berwarna biru. Dapat dilihat Tabel 6.

Tabel 7. Hasil *Testing* Kedua

Jumlah Data testing	Error	Positif Benar (TP)	Positif Salah (FP)	Negatif Salah (FN)
30	5	21	4	1

Pada tabel di atas data dengan perbandingan 80% : 20% terdapat error sebanyak 5 data, positif benar 21, positif salah 4 dan negatif salah 1.

Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasi secara benar dengan total sample data testing yang diuji.

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{Data Testing} - \text{Error}}{\text{Jumlah Data testing}} \\
 &= \frac{60 - 5}{60} \\
 &= \frac{55}{60} \\
 &= 0.91666667 \\
 &= 91.67\%
 \end{aligned}$$

Precision merupakan jumlah sample berkategori positif diklasifikasi benar dibagi dengan total sampel yang diklasifikasi sebagai sample positif. Menghitung nilai *precision* dihitung dengan cara membagi jumlah data benar yang bernilai positif (*True Positive*) dibagi dengan jumlah data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai positif (*False Negative*).

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP} \\
 &= \frac{21}{21 + 4} \\
 &= \frac{21}{25} \\
 &= 0.84 \\
 &= 84\%
 \end{aligned}$$

Recall merupakan jumlah sampel diklasifikasi positif dibagi total sampel dalam testing set

berkategori positif. *Recall* dihitung dengan cara membagi data benar yang bernilai positif (*True Positive*) dengan hasil penjumlahan dari data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai negatif (*False Negative*).

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} \\
 &= \frac{42}{42 + 4} \\
 &= 0.95454545 \\
 &= 95\%
 \end{aligned}$$

F-Measure merupakan *harmonic mean* dari *precision* dan *recall*. Nilai *F-Measure* didapat dari perhitungan pembagian hasil dari perkalian *precision* dan *recall* dengan hasil penjumlahan *precision* dan *recall*, kemudian dikalikan dua.

$$\begin{aligned}
 F\text{-Measure} &= \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \\
 &= 0.89361701 \\
 &= 90\%
 \end{aligned}$$

Tabel 8.Evaluasi dan Validasi Kedua

Presentasi Data	Data Training	Data Testing	Akurasi	Precision	Recall	F-Measure
80% : 20%	120	30	83%	84%	95%	90%

Pada tabel 4 data dengan perbandingan 80% : 20% memiliki nilai akurasi sebesar 83%, *recall* sebesar 95%, *precision* 84% dan *f-measure* 90%.

3.2. Analisa Hasil Pengujian

Pengujian pertama dan kedua. menghasilkan akurasi, *precision*, *recall* dan *f-measure* sebagai berikut:

Tabel 9.Hasil Pengujian

Persentase Data	Data Training	Data Testing	Akurasi	Precision	Recall	F-Measure
60% : 40%	90	60	78%	82%	91%	87%
80% : 20%	120	30	83%	84%	95%	90%

Tabel 8 menunjukkan bahwa nilai *precision* untuk data dengan perbandingan *data training* dan *data testing* 60% : 40% adalah *precision* 82% dan *recall* 91%. Data perbandingan 80% : 20% memiliki nilai *precision* 84% dan *recall* 95%. Data dengan perbandingan 60% : 40% memiliki nilai *F-Measure* 87%, sedangkan data dengan perbandingan 80% : 20% yaitu sebesar 90%.

4. Kesimpulan

Berdasarkan hasil pengujian metode *decision tree* terhadap data pelanggan atau nasabah dapat disimpulkan:

1. Penerapan metode *C4.5* menggunakan data pelanggan atau nasabah dapat digunakan untuk menentukan kepuasan pelanggan atau nasabah.
2. *Ratio data training* yang digunakan mempengaruhi nilai akurasi pada setiap percobaan. Pada percobaan ke-1 nilai akurasi adalah 78% dengan *data training* 60% dan *data testing* 40%. Pada percobaan ke-2 nilai akurasi adalah 83% dengan *data training* 80% dan *data testing* 20%. Dengan hasil ini angka pengujian yang terbaik yaitu dengan akurasi 83% pada percobaan ke-2.
3. Dari percobaan yang telah dilakukan, maka dapat diketahui bahwa setiap percobaan dapat dikatakan

baik, karena sudah terlihat jelas bahwa nilai akurasi yang terus bertambah dan semakin akurat.

5.SARAN

Adapun saran yang disampaikan berdasarkan hasil pengamatan dan analisa selama melakukan penelitian data kepuasan pelanggan adalah sebagai berikut:

- a. Penelitian selanjutnya hendaknya menggunakan data yang lebih banyak agar menghasilkan *rules* yang lebih akurat.
- b. Pengujian metode ini sudah di implementasi pada aplikasi, tetapi perlu dikembangkan kembali agar menjadi aplikasi yang lebih baik.

DAFTAR PUSTAKA

- [1] Elga Rivkyana, "Analisis Tingkat Kepuasan Nasabah Terhadap Kualitas Pelayanan Jasa Pada PT. BANK TABUNGAN NEGARA (PERSERO) TBK Cabang Surakarta," Universitas Sebelas Maret, 2010.
- [2] DANAMON, "www.danamon.co.id"/articles
- [3] Vina Mandasari, "Analisis Kepuasan Konsumen Terhadap Restoran Cepat Saji Melalui Pendekatan Data Mining," Universitas Sriwijaya, Januari 2011.
- [4] Abdul Rohman, "Penerpan Algoritma C4.5 Berbasis Adaboost Untuk Prediksi Penyakit Jantung," Universitas Pandanaran, Agustus 2008.
- [5] Sri Lestari, "Model Klasifikasi Kinerja dan Seleksi Dosen Berprestasi Dengan Algoritma C.45," Institut Informatika dan Bisnis Darmajaya, Desember 2014.
- [6] Kusri, "Algoritma Data Mining," Yogyakarta, Andi, 2009.
- [7] Herlawati, Rahmadya Trias Handayanto, Probowo Pudjo Widodo, "Penerapan Data Mining Dengan Matlab," Rekayasa Sains, Bandung, 2013.
- [8] Prasetyo Eko, "Mengolah Data Menjadi Informasi Menggunakan Matlab," Andi, Yogyakarta, 2014.